Language

What is language for?

- Communication
- But not JUST about communication
  - emotional expression (“Oh #&@!”)
  - thought
  - identity

- And different from animal communication systems

Characteristics of language

Language is . . .

- *Arbitrary:* made up of learned symbols
- *Structured:* governed by a system of rules
- *Generative:* expresses limitless meanings

Characteristics of language

Language is . . .

- *Arbitrary:* made up of learned symbols
- *Structured:* governed by a system of rules
- *Generative:* expresses limitless meanings

apple, manzana, pomme, yabloko

Characteristics of language

Language is . . .

- *Arbitrary:* made up of learned symbols
- *Structured:* governed by a system of rules
- *Generative:* expresses limitless meanings

~~Apple red is.~~ ~~ppola~~

Noun Adjective Verb

Characteristics of language

Language is . . .

- *Arbitrary:* made up of learned symbols
- *Structured:* governed by a system of rules
- *Generative:* expresses limitless meanings

- Possible English sentences with
  …less than 20 words: 10^{30}
  …more than 20 words: infinite!

Why language is amazing:

- Amazingly DIFFICULT:
  - No other organism we know of has language or has been fully successful at learning language.
  - Yet young humans readily learn language and we use it rapidly and effortlessly throughout our lives.

Why language is amazing:

- What makes it so difficult?
  - Language has many intertwined levels.

What it means to know a language

- Phonology – sounds
- Morphology – parts of words
- Syntax – structure
- Semantics – meaning
- Pragmatics – implications

Language rules

Phonology:

- Language sounds and how to combine them.
  - English has 46 phonemes.
  - Other languages have a different number.
  - 200 different phonemes world-wide.

- Certain combinations are “allowed”:
  - snart . . . okay
  - tnart . . . not okay
  - Sbarro . . . okay in Italian, sounds funny in English

Language rules

Phonology:

- Know lots of “rules” we are not conscious of:
  - plural of cabbage?
  - plural of lip?
  - Note the differences in the sounds you just produced:
    - Cabbage\z\
    - Lip\s\
  - These phonological representations vary across languages and across dialects

---

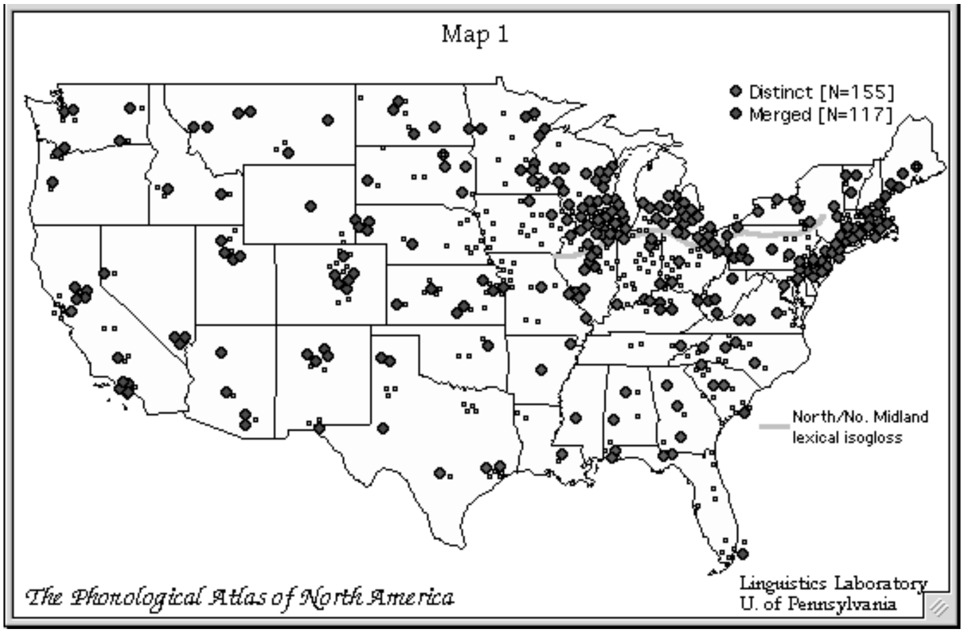
Global variation

- Across languages: thousands of possible phonemes
  - English: ~40 phonemes
  - Polynesian: ~11 phonemes
  - Khoisan: ~141 phonemes
- Some phonemes are more common across languages than others:
  - /m/, /n/, /t/, /d/, /k/, /g/, /s/, /z/

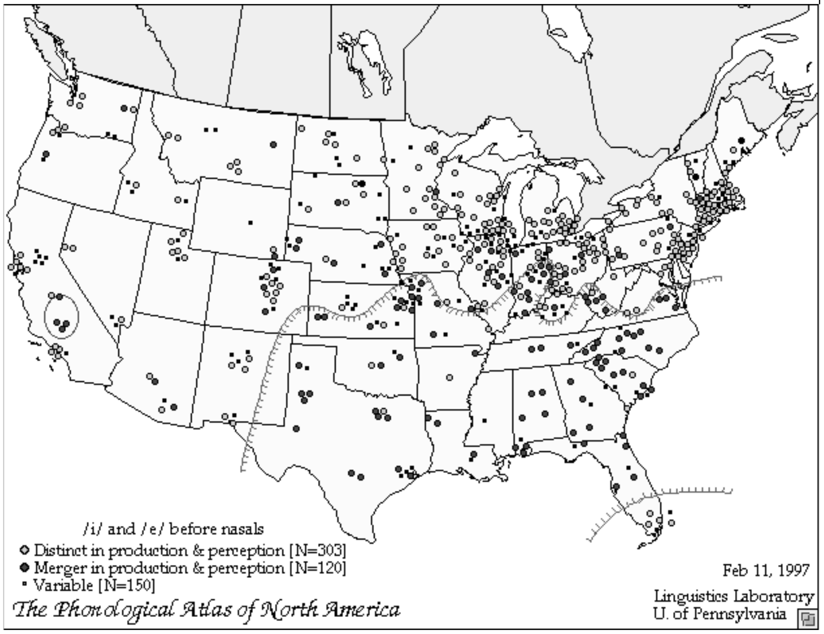
Regional variation

- Do you say these words the same or differently?
  - cot vs. caught?
  - pin vs. pen?
  - Don vs. dawn?
  - heel vs. hill
  - berry vs. Barry vs. bury?
- How do you say *Cool Whip?*

Map of the caught-cot merger



Map of the pin - pen merger


/i/ and /e/ before nasals
○ Distinct in production & perception [N=303]
● Merger in production & perception [N=120]
▪ Variable [N=150]
The Phonological Atlas of North America
Feb 11, 1997
Linguistics Laboratory
U. of Pennsylvania

Northern Cities Vowel Shift

- This is an on-going shift in how vowels are pronounced. It primarily affects speakers from large northern industrial cities (Rochester, Detroit, Chicago..)
- The shift began with the vowel /ae/ (cad) changing to /ie/ (idea)
  - Test yourself: Say *caddyshack*.
  - Are the vowels the same as mine or different?
- This shift prompted several other shifts in the vowel space:
  - /o/ in *cod* shifts to *cad*
  - /oh/ in *cawed* shifts to *cod*
  - /e/ in *ked* shifts to *cud*
  - *cud* shifts to *cawed*
  - /i/ in *kid* moves back to /e/

Language rules

Morphology:

- Parts of words that have “meaning”
  - cat . . . cats
  - walk . . . walked

- We know how to do this at a general level:
  - what is the plural of ferfel?
  - . . . ferfels
  - what is the past tense of blark?
  - . . . blarked

Language rules

Syntax (grammar)

- Ways that words can (and can’t) be put together

This sentence no verb.

This sentence has contains two verbs.

This is not a complete.

This either.

This sentence are not okay.

Language rules

Semantics and pragmatics

- Creating meaning
- Semantics: understanding what words and sentences *refer* to
  - Relatively easy for some kinds of things
  - But much more difficult for others
    - Unicorn
    - Justice
    - The
- Pragmatics: understanding what people mean by what they say
  - Can you pass the salt?

Why language is amazing:

- What makes it so difficult?
  - Language has many intertwined levels.
  - Lots of ambiguity.

Ambiguity

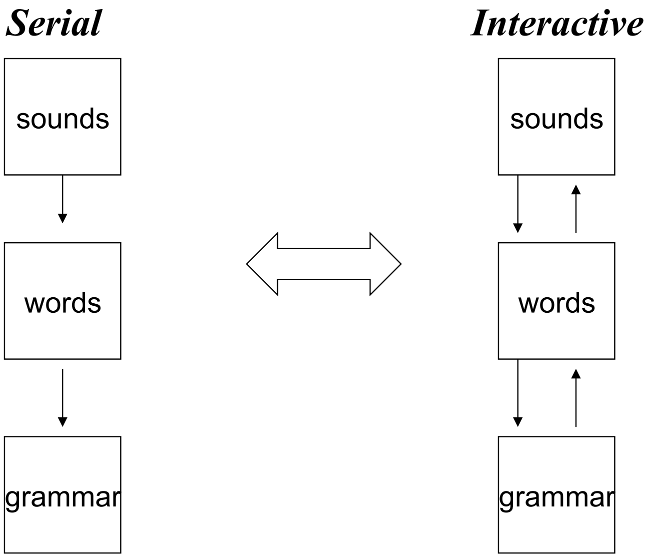
- Understanding language requires computing representations at multiple levels
  - sound, structure, meaning…
- **In all cases**, there is more than one possibility for decoding.
- How (and when) do we choose?

Two solutions to ambiguity

- **Serial bottom-up solution**
  - we process language using default rules (heuristics) that get things right most of the time
  - a separate stage of processing corrects errors using context and other non-linguistic information sources
- **Interactive solution**
  - language is always processed using multiple sources of available information, including context and non-linguistic information

Dealing with ambiguity

Models of Word Recognition:



Serial

sounds

words

grammar

Interactive

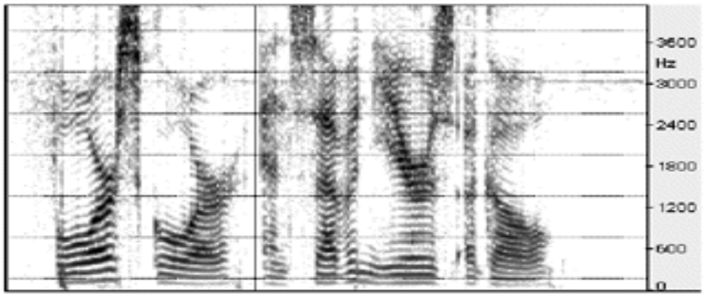
sounds

words

grammar

Ambiguity in speech

- Absence of clear boundaries



"four score and seven years ago"

- Multiple ways to divide up sounds into words

Top-down processing

- Phonemic restoration effect (Warren):
  - People hear sentence where sound has been replaced by a cough.
    - “The state governors met with their respective legi*latures convening in a capital city.”
  - 19/20 subjects said no sound was missing.
  - None of the subjects could accurately locate the cough.

Gap

Normal

Top-down processing

- (More) phonemic restoration effect (Warren):
  - Subjects heard sentences like:
    - It was found that the *eel was on the axle.
    - It was found that the *eel was on the shoe.
    - It was found that the *eel was on the orange.
  - Again, didn’t notice missing sound.
  - Heard “wheel”, “heel”, and “peel” respectively.

Cohort Model (Marslen-Wilson & Tyler)

- Words consistent with input become active
  - *Cohort:* set of words consistent with first syllable
- Words in the cohort are eliminated when they become inconsistent with input.
- Words also eliminated due to contextual incongruity.
- Processing ends when there is one word left in the cohort.

/ka/

cat captain catch capitalism

/kap/

captain capitalism

Communism is not the same as /kap/

capitalism

Marslen-Wilson & Tyler

Monitor for target words

- Normal

The church was broken into last night.

Some thieves stole most of the gold off the altar.

- Syntactic

The power was located in green water.

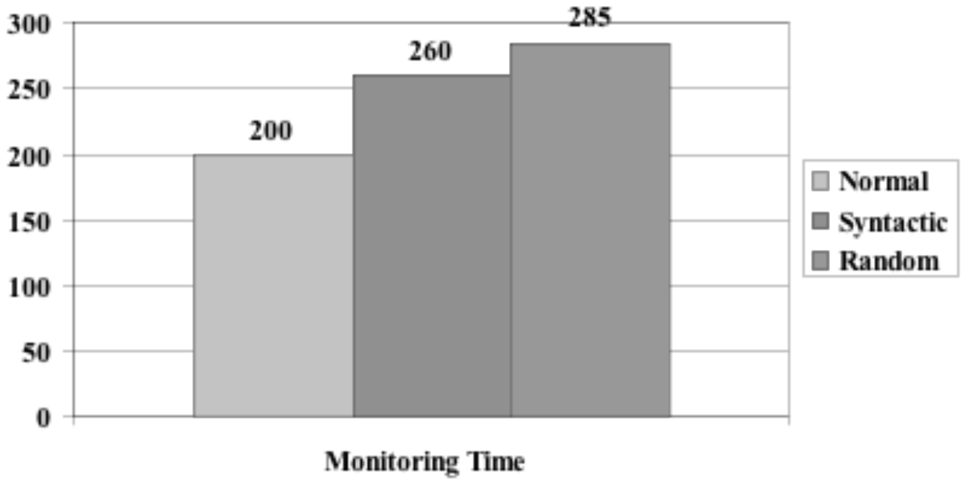
No buns puzzle some in the gold off the text.

- Random

In was great power water the located.

Some the no puzzle buns in gold text the off.

Marslen-Wilson & Tyler



Interim summary

- Lots of ambiguity in the speech stream.
- Seem to use knowledge of words and their meanings to disambiguate.
- Key question is *when* and *how* this top-down information is used.
- Cohort model suggests that sensory analysis and context information are both used (but independently).

Word meaning ambiguity

- Identifying words not enough.
- Many words have multiple meanings:

- *I have to say that I think Pam's real talent is getting wasted at her current job. She always seems to be loaded with work to do. Yet, she still takes the time to go through several drafts before she thinks her papers are ready for submission. Indeed, I believe that the number of papers she has produced (while staggering), does not accurately reflect her true ability. Thus, I would make sure there are no bars in the way of this woman coming to work for you.*

Word meaning ambiguity

- Many words have multiple meanings:
  - bank
    - river or financial institution?
  - train
    - vehicle or bridal gown part?

Word meaning ambiguity

- How do we rapidly pick the right one?
- Can use frequency:
  - more likely to encounter “train” as vehicle
- Can use context:
  - “The maid of honor smoothed the bride’s train.”
  - “The business traveler decided to go by train.”
- Or maybe DON’T pick …

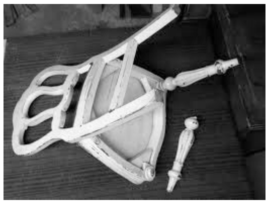
Word meaning ambiguity

- How can we test?
- Priming paradigm:
  - The maid of honor smoothed the bride’s train.
  - The business traveler decided to go by train.
- After each sentence get a “probe word” that must rapidly respond to:
  - gown
  - engine
  - cat
- Should get priming for any meaning that is active.

Word meaning ambiguity

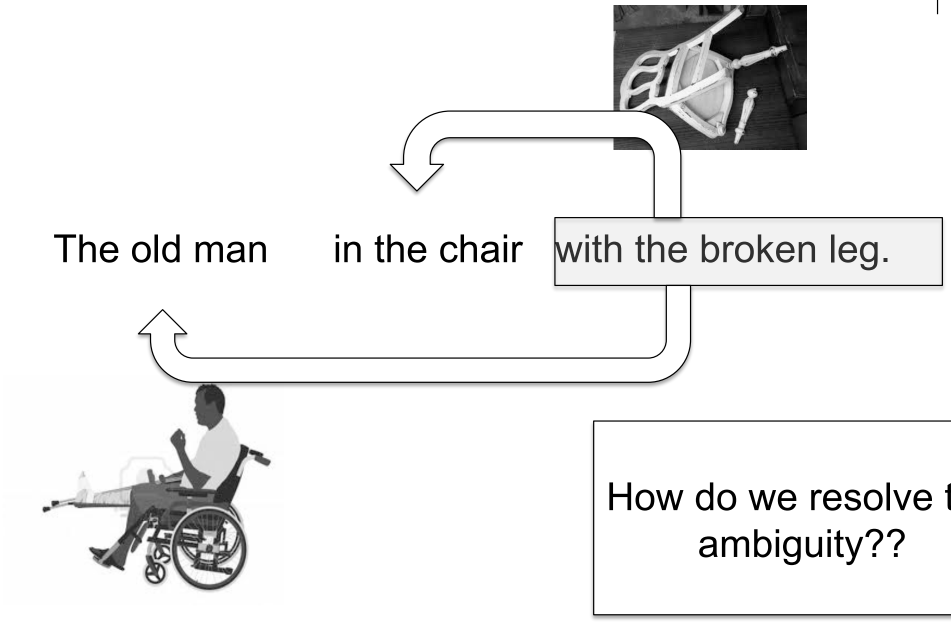
- Results?
- If test **immediately** see priming for all meanings:
  - The maid of honor smoothed the bride’s train.
  - Priming for GOWN and for ENGINE.
- If wait a short time after sentence before giving probe word, get priming only for contextually appropriate meaning:
  - Now priming only for GOWN.
- These results are consistent with a serial, bottom-up view.

Ambiguity beyond words



- Syntactic Ambiguity
  - Sentences or phrases that have two or more interpretations because there are multiple ways to “parse” them.
- **The old man in the chair with the broken leg.**

Syntactic "attachment" ambiguity



The old man in the chair with the broken leg.

How do we resolve the ambiguity??

Two solutions to ambiguity

- **Serial solution**
  - use a heuristic to resolve the ambiguity
  - if the heuristic is wrong, revise your interpretation later
- **Interactive solution**
  - use multiple top-down & bottom-up cues to figure out what the speaker intended

Context effects?

- Example: “landing planes”
  - planes that are landing?
  - the act of landing planes?
- Study by Tyler & Marslen-Wilson
  - If you walk too near the runway, landing planes …
  - If you’ve been trained as a pilot, landing planes …
  - If you walk too near the runway, landing planes ARE
  - If you’ve been trained as a pilot, landing planes IS
- Results show that people expect contextually correct verb (not the same verb each time).

Interim summary II

- In addition to sound ambiguities, have ambiguities in word meaning and the structure and meaning of phrases and sentences.
- Evidence for both bottom-up processing and top-down processing:
  - **bottom-up**: early activation of both meanings of an ambiguous word
  - **top-down**: sentence context guides syntactic “parsing”

Discourse processing

- In larger texts, lots of things go “unsaid”.
- People need to make **elaborative inferences** -- filling in missing details
  - He pounded the nail in.
  - *with a hammer?*
  - *with a shoe?*
- Do we make inferences only when we HAVE to (in order to make sense of language) or all the time?

Testing inference-drawing

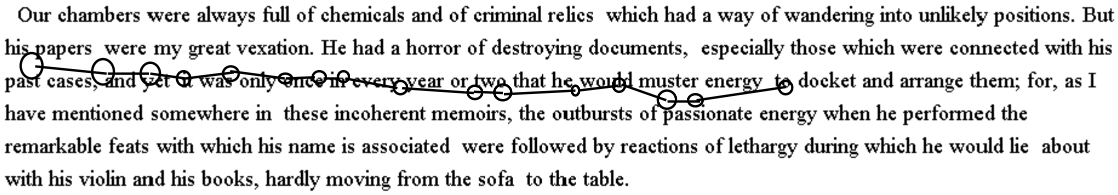
All the mugger wanted was to steal the woman’s money. But when she screamed, he (assaulted/stabbed) her with his weapon in an attempt to quiet her. He looked to see if anyone had seen him. He threw the knife into the bushes, took her money, and ran away.

Question: do people infer that a knife was used when they read the word “stabbed”?

Using eye-movements

- Look at how eyes move when reading.
  - Saccades: “jumps” from one position to another
  - Fixations: times when eyes are still

Using eye-movements



Our chambers were always full of chemicals and of criminal relics which had a way of wandering into unlikely positions. But his papers were my great vexation. He had a horror of destroying documents, especially those which were connected with his past cases, and yet it was only once in every year or two that he would muster energy to docket and arrange them; for, as I have mentioned somewhere in these incoherent memoirs, the outbursts of passionate energy when he performed the remarkable feats with which his name is associated were followed by reactions of lethargy during which he would lie about with his violin and his books, hardly moving from the sofa to the table.

- Can measure amount of time people spend fixating particular words.
- Look less time for predictable words.

Testing inference-drawing

All the mugger wanted was to steal the woman's money. But when she screamed, he (assaulted/stabbed) her with his weapon in an attempt to quiet her. He looked to see if anyone had seen him. He threw the knife into the bushes, took her money, and ran away.

Question: how long do people fixate knife?

Testing inference-drawing

- If not making inferences: same with both verbs
- If making inferences: less time after stabbed

- Results show shorter fixation for "knife" after "stabbed" than after "assaulted."
- Suggests people already activated "knife."

Conclusions

- Get a rapid, ambiguous, under-specified signal.
- Must extract sounds, words, structure, meaning.
- Two approaches.
  - Serial/bottom-up:
    - Analyze signal and apply a set of carefully chosen rules to maximize efficiency.
    - Use context to fix errors.
  - Interactive:
    - Use all the information you have (in signal, in memory, in context) as soon as you can.
    - Have principles for dealing with conflict between information sources.

Question

Consider the following pair of sentences:

The cop shot the spy with the gun.

The cop arrested the spy with the stolen documents.

What does a modular, bottom-up model predict about how these sentences should be parsed (interpreted) and what behavioral consequences that should have?